

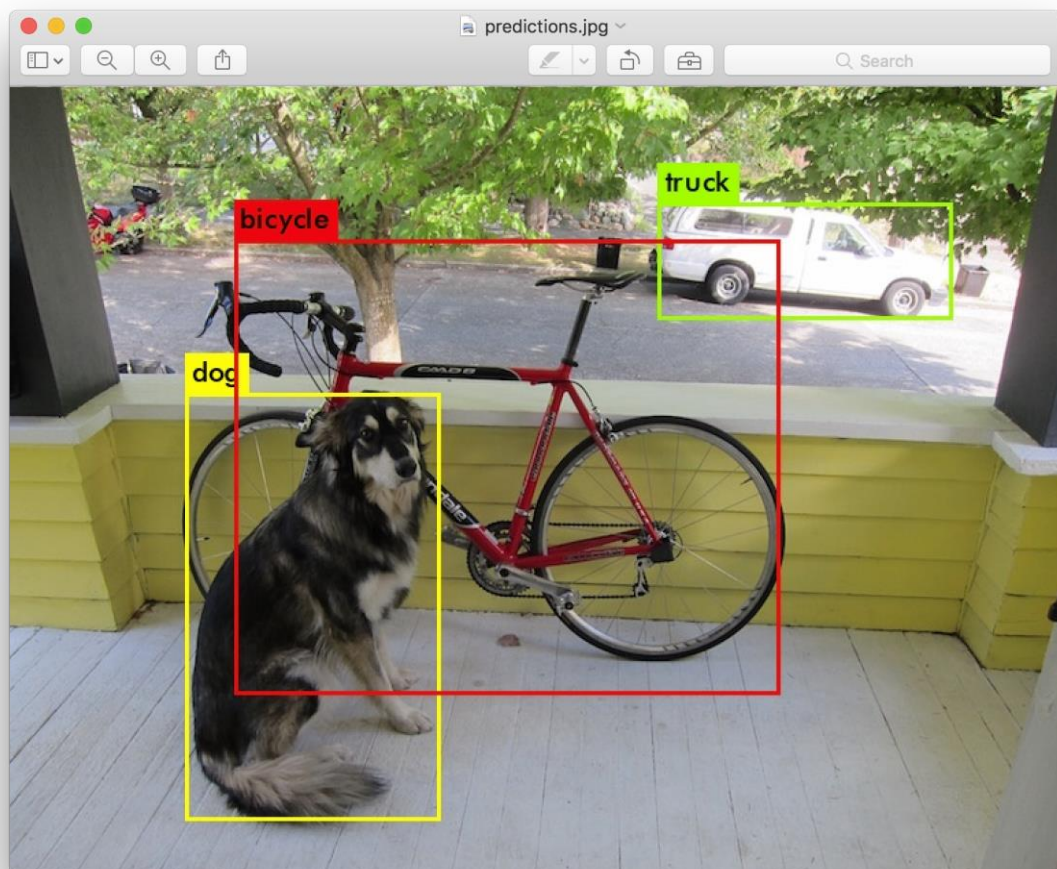
Report về object detection

Thành viên nhóm:

Võ Huy Khôi	18520949
Hoàng Xuân Thắng	18521391

Object detection là một bài toán trong computer vision nhằm định vị đối tượng và nhận dạng đối tượng đó trong một bức ảnh. Object detection là sự kết hợp giữa image classification và object localization:

- Image classification: phân loại, gán nhãn cho hình ảnh.
 - Object localization: định vị đối tượng cần gán nhãn bằng cách vẽ một đường bounding box xung quanh đối tượng.
- ⇒ Output đối với object detection: xác suất dự đoán cho đối tượng trong ảnh và vị trí của bounding box: chiều dài b_h , chiều rộng b_w , tọa độ điểm trung tâm bounding box (b_x , b_y).



Hướng xử lý: sử dụng phương pháp sliding window với các shape khác nhau. Trượt một cửa sổ qua toàn bộ tấm ảnh, thực hiện predict trên từng cửa sổ và trả về label nếu tìm thấy đối tượng.

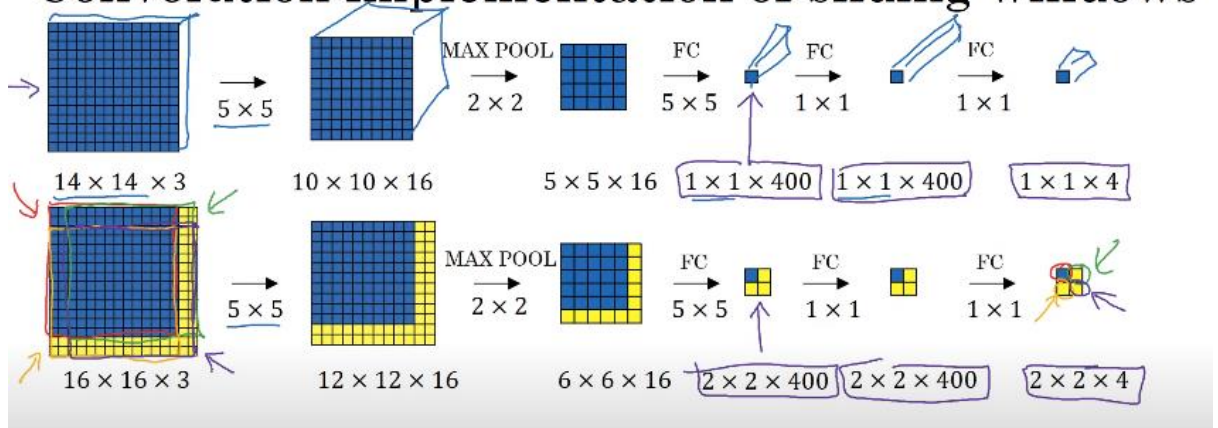
Sliding windows detection



Tuy nhiên cách làm này rất tốn tài nguyên tính toán.

⇒ Khắc phục bằng phương pháp convolutional sliding window: chuyển lớp FC cuối thành các lớp Conv và sử dụng filter 1×1 với softmax activation. Với input shape = (i_height, i_width) ; window shape = (w_height, w_width) -> output có kích thước $((i_height - w_height), (i_width - w_width), num_class)$. Với mỗi giá trị trong ma trận output[:, :] sẽ tương ứng với slide window ban đầu -> xác định được vị trí của vật thể.

Convolution implementation of sliding windows



Landmark detection:

Tìm tọa độ (l_x, l_y) các điểm cho các đặc trưng trong ảnh. Ứng dụng trong các bài toán nhận diện khuôn mặt dựa trên landmark của mắt, môi, miệng,... và nhận diện dáng người dựa trên landmark của vai, lưng,...



Đánh giá mô hình object detection:

Sử dụng 2 hàm loss cho 2 nhiệm vụ của object detection:

- Image classification: cross-entropy loss đối với bài toán multi label classification, logistic loss đối với bài toán binary classification.
- Object localization: localization loss tính giá trị mất mát khi dự đoán bounding box

Cross-entropy loss:

Trong các bài toán multi classification thì label sẽ được biểu diễn dưới dạng một one-hot vector. Tức là khi đó mỗi output sẽ không còn là một giá trị tương ứng với mỗi class nữa mà sẽ là một vector có đúng 1 phần tử bằng 1, các phần tử còn lại bằng 0. Phần tử bằng 1 nằm ở vị trí tương ứng với class đó, thể hiện rằng điểm dữ liệu đang xét rơi vào class này với xác suất bằng 1.

Khi sử dụng mô hình Softmax Regression, với mỗi đầu vào x , ta sẽ có *đầu ra dự đoán* là $a = \text{softmax}(Wx)$. Trong khi đó, *đầu ra thực sự* chúng ta có là vector y được biểu diễn dưới dạng one-hot coding. Để tính loss cho dự đoán a và giá trị đầu ra thực sự người ta có thể sử dụng Cross Entropy Loss.

$$CE = - \sum_i^C t_i \log(s_i)$$

Trong đó: i : số class

t_i : label của class i (1 hoặc 0)

s_i : xác suất được dự đoán của mô hình cho đối tượng

Logistic loss:

$$l(z) = -\log \left(\prod_i^m \mathbb{P}(y_i|z_i) \right) = - \sum_i^m \log (\mathbb{P}(y_i|z_i)) = \sum_i^m -y_i z_i + \log(1 + e^{z_i})$$

Trong đó: y_i : label thực tế

Z_i : label dự đoán của mô hình

Hàm loss trong object detection:

$$\mathcal{L}_{\text{loc}} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2]$$

$$\mathcal{L}_{\text{cls}} = \sum_{i=0}^{S^2} \sum_{j=0}^B (1_{ij}^{\text{obj}} + \lambda_{\text{noobj}}(1 - 1_{ij}^{\text{obj}}))(C_{ij} - \hat{C}_{ij})^2 + \sum_{i=0}^{S^2} \sum_{c \in \mathcal{C}} 1_i^{\text{obj}} (p_i(c) - \hat{p}_i(c))^2$$

$$\mathcal{L} = \mathcal{L}_{\text{loc}} + \mathcal{L}_{\text{cls}}$$

Trong đó:

- 1_i^{obj} : An indicator function of whether the cell i contains an object.
- 1_{ij}^{obj} : It indicates whether the j -th bounding box of the cell i is “responsible” for the object prediction (see Fig. 3).
- C_{ij} : The confidence score of cell i , $\text{Pr}(\text{containing an object}) * \text{IoU}(\text{pred}, \text{truth})$.
- \hat{C}_{ij} : The predicted confidence score.
- \mathcal{C} : The set of all classes.
- $p_i(c)$: The conditional probability of whether cell i contains an object of class $c \in \mathcal{C}$.
- $\hat{p}_i(c)$: The predicted conditional class probability.

Intersection over union: để đánh giá cần:

- Những ground-truth bounding box (bounding box đúng của đối tượng, ví dụ như bounding box của đối tượng được khoanh vùng và đánh nhãn bằng tay sử dụng trong tập test.)
- Những predicted bounding box được model sinh ra.

IoU được tính bằng công thức: area overlap/ area union

- Area overlap: số pixel trùng nhau giữa predicted bounding box và ground-truth bounding box.
- Area union: tổng số pixel giữa ground-true và predicted bounding box



Intersection over Union (IoU)

$$= \frac{\text{size of } \text{[yellow box]}}{\text{size of } \text{[green box]}}$$

“Correct” if $\text{IoU} \geq 0.5$

Trong object detection, một đối tượng thường có thể có nhiều predicted bounding box.

⇒ Hướng giải quyết:

- Loại bỏ tất cả bounding box với xác suất dự đoán đối tượng < 0.6
- Với các bounding box còn lại: chọn bounding box có xác suất dự đoán cao nhất và loại bỏ những bounding box có $\text{IoU} > 0.5$ với bounding box được chọn vì những bounding box này đang khoanh vùng cùng 1 đối tượng đối với bounding box được chọn.